DOCUMENT RESUME

ED 148 107 FL 009 003

AUTHOR Stansfield, Charles
TITLE The Cloze Procedure as a Progress Test.
PUB DATE 21 Oct 77
NOTE 11p.; Paper presented at the annual Rocky Mountain Modern Language Association meeting (thirty-first, Las Vegas, Nevada, Oct. 21, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; *Cloze Procedure; College Language Programs; Comparative Testing; Cultural Education; Educational Research; Higher Education; *Language Instruction; *Language Tests; *Multiple Choice Tests; Reading Ability; *Reading Tests; Second Language Learning; *Spanish; Testing; Test Results

ABSTRACT

This paper reports on a pilot study conducted to determine the possible use of the cloze procedure as a substitute measure of achievement in a second-year Spanish culture and civilization class. Twenty students enrolled in a third-semester Spanish class at the University of Colorado were simultaneously given multiple choice tests and cloze tests on four cultural essays they studied during the course. Correlations between the two instruments were highly significant and ranged from .32 to .67. It is concluded that the cloze test is a better discriminator of the acquisition of cultural knowledge and is much easier to construct than a multiple choice reading test. Other advantages and disadvantages of using the cloze procedure in the classroom are discussed. (Author/AM)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED148107

"The Cloze Procedure as a Progress Test"*

During the past 25 years much research has been done on the cloze procedure, or cloze test as it is more often called in the field of foreign language education. Taylor was the first to experiment with it, when in 1953, he studied it as a measure of contextual redundancy in his dissertation at the University of Illinois.[1] Taylor named the technique the cloze procedure, deriving it from the concept of closure in Gestalt psychology. Gestaltists believe that learning follows a sequence through which one first understands the whole, or broader issues, and then grasps the individual details. Applied to visual perception, the theory suggests that in looking at a tree, we first see it as a single unit, rather than as a collection of individual leaves and branches. Once this whole is perceived, we can then begin to make note of individual details such as groupings of foliage or missing branches. Closure refers to the natural tendancy to mentally close the gap or fill in the spaces, such as in perceiving a not-quite-complete circle as a whole circle. Likewise, the cloze procedure requires the student to perceive the whole, by filling in missing words, as if they were not missing at all.

The cloze procedure was originally used as a measure of the readability or difficulty of a passage of prose. It is constructed by systematically eliminating every Nth word in a passage with the distance between deletions usually being

*Paper presented at the 31st Annual Meeting of the Rocky Mtn. Modern Language Association, Oct. 21, 1977, Las Vegas, Nevada.

F1009003

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Charles Stansfield

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPPESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

neither less than five words nor more than ten words. The examinee attempts to reconstruct the passage by filling in the missing words. Taylor later discovered the cloze could serve as a measure of prior familiarity with, or mastery of a specific body of knowledge. In a study supported by the U.S. Air Force, he administered an 80-item cloze test and a multiple choice comprehension test on the Air Force system of supply to three groups of servicemen. One group (N=52) received an easy cloze test, based only on verb auxilaries, conjunctions, pronouns, and articles. Another group (N=48) received a cloze test of moderate difficulty based on the systematic deletion of every fifth word. A third group (N=52) received a hard cloze test which deleted only adverbs, verbs, and nouns. Only those fill-ins which matched the original word were counted as right. Following this pretest, the subjects were asked to study a 3200 word technical article on the Air Force supply system. One week later the subjects were retested. The pretest served as a measure of prior knowledge and the post-test served as a measure of knowledge acquired following instruction. While all three groups showed significant gains on the post-test, scores during both administrations correlated highly, with Pearson coefficients ranging from .58 to .92. Taylor concluded that cloze scores are influenced by pertinent specific knowledge and overall language facility.[2]

Between 1957 and 1959, Carroll, Carton, and Wilds con-

ducted a feasibility study on the possible inclusion of the cloze procedure in the College Entrance Examination Board's foreign language tests. It may surprise researchers today to learn that the authors rejected the cloze procedure as a proficiency measure, "because they are relatively unreliable and too heavily affected by various sources of extraneous variance"[3]. While not always clear on this matter, the authors indicate the sources of extraneous variance are the examinee's level of general knowledge, familiarity with the subject, ability to assimilate new information while completing the passage, and general reasoning ability. Perhaps because of this conclusion, and because of the instrument's lack of face validity, a full decade passed before another published study directed attention to the cloze procedure in the assessment of language proficiency.

In 1968, Darnell correlated scores obtained with a modified cloze procedure called clozentropy, with scores obtained on the Test of English as a Foreign Language published by Educational Testing Service. Using 48 foreign students at the University of Colorado as subjects, he obtained a correlation of .84 between the two instruments.[4]

Following publication of Darnell's findings, Bowen,[5] Kaplan and Jones,[6] Oller and Conrad,[7] Stubbs and Tucker,[8] and many others investigated and confirmed the utility of this instrument in assessing overall second language proficiency. Other uses are yet to be discovered.

The present study was intended as a pilot project and presents the results of an informal classroom experiment conducted by the author to investigate the possible application of the cloze procedure as a surrogate measure of achievement in second year reading courses in a foreign language. The author wanted to ascertain whether the degree of familiarity with a reading selection, would determine a student's score on a cloze test. Because of Taylor's previous finding, it was hypothesized that there would be a high correlation between the scores obtained on multiple choice reading tests and cloze passages taken from the reading selections studied by the class.

The following sentence illustrates how it was anticipated the process would work:

Once _____ a time there were _____ little pigs.
A B

While the native speaker of English could probably write "upon" on line A, he would not know that the answer to B was "three" unless he was familiar with the story. Otherwise, he could be expected to fill in such words as "some", "these", or "several". A foreign language learner, lacking the native speaker's structural proficiency, might miss both words, unless very familiar with the story.

Method. During the summer of 1975, the author taught a third semester Spanish course to 20 students at the University of Colorado. The curriculum was based on the preliminary edition of a review grammar and collection of cul-

tural essays by Copeland, Kite, and Sandstedt.[9] The cultural essays were oriented toward the social sciences and dealt with various aspects of the Hispanic world such as family, death, religion, pre-Columbian civilizations, and economic problems. Each selection was approximately 2000 words in length.

After reading, discussing, and studying each essay, the students were given a 15-item, four-option multiple choice test based on the content of the selection. Immediately afterwards, they were given two paragraphs selected from the passage and systematically mutilated according to the cloze procedure. The two paragraphs selected were from different parts of the essay and contained a total of 30 deletions for students to fill in. Following each testing session, the two sets of scores were correlated (Pearson's r) on a hand calculator. The obtained coefficients are listed below.

TABLE I

Product Moment Correlations Between Cloze and Multiple Choice Scores on Four Tests.

Test	Correlation	Nth Word Deleted
I	.32	5th
II	.67	7th**
III	.61	8th*
IV	.62	9th*

*$p < .005$
**$p < .001$

Results and Analyses. On the first cloze test, Taylor's scheme of deleting every fifth word was followed. This resulted in a cellar effect in that the highest score was only 18 of 30 items correct. This attentuated distribution would account for the low correlation between scores on this administration, since a reduction in variance will always lower the correlation between scores that are less than perfectly reliable. On the second test every seventh word was deleted. While this improved the distribution, most cloze scores were still less than 50% correct. On the third and fourth tests, the number of words between each deletion was again increased until on the fourth test, the mean score was 68% correct. Mean scores on the four multiple choice tests probably prevented the correlation from reaching higher levels throughout the course. Nonetheless, the correlations on tests II, III, and IV were highly significant.

Discussion. It is interesting to ponder the implications of this pilot study. Although the correlations were significant, the common variance can not be attributed to knowledge of the content alone. It could be that a general language proficiency factor is responsible for similar results on both tests. Undoubtedly, this factor would always be responsible for some similarity in scores. It would be interesting for future studies to include an unfamiliar cloze passage as a general proficiency control measure and then compare correlations. If the familiar

cloze correlated better with the unfamiliar cloze than with the multiple-choice test, then it would have little usefulness as a progress test. On the other hand, if its correlation with the multiple-choice progress test were substantially greater, then the cloze could serve as an easy-to-construct substitute measure.

Another question which merits the attention of future research on this topic is the effect of selective rather than systematic deletions. Taylor found that function words (articles, pronouns, auxilaries, and conjunctions) were much easier to replace than major content words (nouns and verbs), although the correlations obtained with separate cloze tests of each did not exceed that obtained by systematically deleting every fifth word.[10] Nonetheless, the exclusion of function words from the passage mutilation process could produce a better test of knowledge acquisition for foreign languages since it might partially control for differences in language proficiency among learners. As Oller demonstrated in his study comparing exact word and contextually acceptable scoring methods, the findings of cloze research based on native speaking subjects can not be generalized to second language learners.[11]

Use of the cloze test in the classroom situation resulted in certain reactions which merit reporting. In general, students disliked it and felt that the instructor was testing the remembering of words instead of the acquisition of cultural knowledge. Also, since the results of

the cloze test were counted in determining the final grade, students studied the essay assiduously until they not only knew its content, but could remember in what order ideas were discussed. As a result, the multiple choice reading tests became very easy, in spite of an attempt to make them demanding, and this prevented the obtaining of a reliable spread of scores. On the other hand, the instructor was pressured to make the cloze test easier by a barrage of initial complaints. Therefore, on successive tests the distance between deletions was increased. Once the scores improved, complaints about the cloze subsided.

Use of the cloze test also affected students' retention of vocabulary. Following Taylor's procedure, only the exact word which occurred in the text was accepted. As a result, students made a greater than normal effort to absorb new words into their active vocabulary so that they could produce them on the cloze test. The same effect was also noticed for structure. The cloze passages forced students to use many structural items which would normally be above the productive syntactic abilities of second-year students. These included contextually appropriate relative pronouns, conjunctions, miscellaneous clause and sentence relators, and compound verb tenses. As a result, the use of the cloze had a reinforcing effect on the grammar review and it accelerated language acquisition in general. This is because the cloze prohibited the students from becoming sat-

isfied with learning just the content of the passage. It made them acquire a more sophisticated level of language. Instead of permitting a disparity between receptive and productive skills, it forced the student to integrate these into a single active system.

In summary, in my first experience with the cloze procedure as a progress test, it performed better in its teaching than its testing function. While adverse student attitudes were initially a problem, these were overcome with experience and practice. While the cloze did not show high correlations with the multiple choice test of content knowledge, this was probably due to the latter's lack of discrimination power. The author felt that the cloze was functioned better as a measure of familiarity with the content, while it encouraged students to improve their overall proficiency. Also, it was much easier to construct and just as easy to grade. Future studies, in both controlled experimental and classroom settings, will have to verify the procedure's utility as a progress test. For the present however, these results seem encouraging.

Charles Stansfield
University of Colorado
Boulder

NOTES

[1]Wilson L. Taylor, "Cloze Procedure: A New Tool for Measuring Readability", Journalism Quarterly XXX (Fall, 1953), 414-438.

[2]Wilson L. Taylor, "Cloze" Readability Scores as Indices of Individual Differences in Comprehension and Aptitude." Journal of Applied Psychology, XLI (April, 1957), 19-26.

[3]John B. Carroll, Aaron S. Carton, and Claudia Wilds, "An Investigation of "cloze" items in the Measurement of Achievement in Foreign Languages." ERIC ED 021 513, p. 116.

[4]Donald K. Darnell, "The development of an English Language Proficiency Test of Foreign Students Using a Clozentropy Procedure." ERIC ED 024 039.

[5]Donald K. Darnell, "Clozentropy: A Procedure for Testing English Language Proficiency of Foreign Students", Speech Monographs, XXXVII (March, 1970), 36-46.

[5]Donald J. Brown, "A Tentative Measure of the Relative Control of English and Amharic by Eleventh Grade Ethiopian Students", Workpapers in TESL II, (June, 1969), 69-89.

[6]R. Kaplan and R.A. Jones, "A Cloze-Procedure Test of Listening for University Level Students of ESL." Unpublished paper, University of Southern California, 1970.

[7]John W. Oller Jr. and Christine A. Conrad, "The Cloze Procedure and ESL Proficiency." Language Learning, XXII, (June, 1972), 1-15.

[8]Joseph Bartow Stubbs and G. Richard Tucker, "The Cloze Test as a Measure of English Proficiency." The Modern Language Journal, LVI (September-October, 1974), 239-241.

[9]John B. Copeland, Ralph Kite, and Lynn Sandstedt, Civilización y cultura. (New York: Holt, Rinehart, and Winston, 1977). And John B. Copeland, Ralph Kite, and Lynn Sandstedt, Conversación y repaso. (New York: Holt, Rinehart, and Winston, 1977).

[10]Wilson L. Taylor, "Cloze Readability Scores as Indices of Individual Differences in Comprehension and Aptitude." op. cit.

[11]John W. Oller Jr. "Scoring Methods and Difficulty Levels for Cloze Tests of ESL Proficiency." The Modern Language Journal, LVI (March, 1972), 151-158.